

A new focus in the science of networks: towards methods for design

BY YAN WAN*, SANDIP ROY AND ALI SABERI

*School of Electrical Engineering and Computer Science, Washington State
University, PO Box 642752, Pullman, WA 99164-2752, USA*

In recent years, a realization that networks are ubiquitous in the natural and engineered worlds has led to a burgeoning interest in finding commonalities in their structures and dynamics. Here, we introduce a new design focus in this science of networks by proposing generic methods for *synthesizing network controllers that exploit the topological structure*. That is, we motivate a canonical controller synthesis problem for networks that has applications in such diverse areas as virus-spreading control and air traffic flow management. We address this design problem by using new techniques from *decentralized control theory*. Specifically, we mesh optimization machinery together with eigenvalue sensitivity and graph theory notions to identify general structural features of optimally actuated networks. From these features, we are in turn able to explicitly construct high-performance controllers, i.e. the ones that best exploit the network's topological structure. Our general approach for controller design is important because it both provides a broad insight into the structure of well-designed networks and contributes engineering solutions in numerous application areas (e.g. reduction in management delays and human-controller workload in air traffic systems).

Keywords: decentralized control; network analysis; epidemic control;
air traffic flow management

1. Introduction

Historically, efforts have been made to model various network's dynamics, including electrical power system transients (Anderson & Fouad 2003), metapopulation dynamics (Levins 1969) and (more recently) biological-oscillator synchronization (Strogatz 1993), among many others. In the last 20 years or so, these individual efforts have evolved into a 'science of networks' (e.g. Bak *et al.* 1987; Albert & Barabási 2000; Watts 2003, 2004; de Menezes & Barabási 2004; Newman *et al.* 2006). That is, scientists and engineers have identified commonalities among the structures and dynamics of many natural and man-made networks. They have thus sought to codify the typical structural features of networks, and in turn to understand how these features modulate the network's dynamics. There have been numerous outcomes of this science of

All authors contributed equally to this work.

* Author for correspondence (ywan@eecs.wsu.edu).

networks, ranging from statistical prediction of failure event sizes (Carreras *et al.* 2002) to conceptual discussions about whether the structures of engineered (designed) and natural networks are similar (Carlson & Doyle 1999). These studies have been buttressed by an impressive body of analytical research on matrix algebra, which proves valuable for network analysis because interactions between network components (parts) can be codified using matrices. Of particular note in this broad domain are works on *D-stability* and *diagonal Lyapunov stability* (which are useful, e.g. in studying population dynamics; Hershkowitz 1992), efforts to characterize pertinent classes of matrices such as *non-negative matrices* (Berman & Plemmons 1994) and the more recently developed field of *spectral graph theory* (Chung 1994).

While the science of networks has been extensively developed, it is our contention that these efforts have largely focused on modelling, i.e. on predicting the network's structural and dynamic characteristics. What has heretofore *not* been addressed in a general way is the critical task of network management and design, i.e. of efficiently using available resources to improve a network's dynamic performance by exploiting its topology. In myriad application areas, addressing such design tasks would permit significant improvement of the network performance (e.g. reduction in air traffic delays or less-intrusive containment of virus spread). Just as networks have commonalities in their structures and dynamics, one would hope that management and design problems for various networks could be abstracted to a common core, and in turn these features of good designs (or of networks upon design) could be identified. Here, we motivate through many examples a canonical problem of network *controller* (or management scheme) synthesis. Through this canonical problem, we make clear that network controller synthesis problems *require new matrix analysis methods*, and hence develop matrix-theoretic tools for synthesis. In this process, we also identify characteristics of well-controlled networks.

In discussing the need for design, it is worth noting that communication network engineers (in domains such as ad hoc networking, Internet congestion control and computer-work elimination, among others) have recently begun contemplating the role played by a network's graph in system design (e.g. Costa *et al.* 2004; Hekmat 2006) and have studied optimization of certain network (routing or control) algorithms (e.g. Kelly 2003; Gevros & Crowcroft 2004; Wang *et al.* 2006). These efforts are meshed with our philosophy that network performance must be shaped/optimized through the use of the network topology. However, the graph-theoretic studies (e.g. Costa *et al.* 2004; Hekmat 2006) largely focus on *evaluating* the impact of the graph structure on the network performance for a particular design, or, at best, pursue the design of a few global graph statistics (e.g. node degree distribution). On the other hand, the heterogeneous design studies are focused on optimizing static (steady-state) performance measures (Wang *et al.* 2006) using, for example, a linear programming formulation, or on optimizing congestion/throughput at single bottlenecks (Kelly 2003; Gevros & Crowcroft 2004). In contrast, we motivate and address the design of network *dynamics* using fine (local) controls, while still making explicit the role played by the graph topology. We believe strongly that our efforts to *control/shape* the network dynamics by assigning local resources to exploit the graph topology are of significant interest in communications applications, in addition to the network applications that we develop here.

The tools that we develop are inspired by problems from the field of decentralized control theory (Wang & Davison 1973; Siljak 1994) and, in fact, contribute to research efforts in this field. In contrast to the complex networks literature, control theory has long focused on *designing* dynamics through feedback. Decentralized control theory—which is concerned with systems whose parts each have incomplete ability to observe and actuate the dynamics—can potentially address network management/control problems. Unfortunately, decentralized controllers historically have been designed to achieve performance goals by making individual system components *robust* to any network impact (Siljak 1994). Such a paradigm is not viable for modern networks, in which individual parts cannot possibly achieve performance aims without using their network connections, and hence modern control strategies *must take advantage of the network topology*. With this requirement in mind, only the seminal work of Wang & Davison and works derived thereof—which give conditions for the existence of *stabilizing* dynamic controllers for a very broad class of decentralized systems—are viable for the control of modern networks (Wang & Davison 1973). In fact, beyond the existence results provided in Fisher & Fuller (1958) and Wang & Davison (1973; which consider *static* or memoryless controllers), very little is known about the complex network control. In particular, novel methods for synthesizing high-performance controllers are badly needed.

Over the last 5 years or so, there has been a recognition that practical decentralized control requires analysing the structure and dynamics of complex networks using graph theory methods (Roy *et al.* 2006; Olfati-Saber *et al.* 2007). This recognition has led to the formulation of a suite of algorithms and controllers for *autonomous agent networks* (e.g. vehicle teams, sensor networks) that use the network's graph topology. In turn, various tests for checking whether such algorithms/controllers can complete desired tasks have been developed. However, systematic tools for designing high-performance controllers have not been developed even for these applications. Our recent work (Roy & Saberi 2007) proposes a methodology for designing high-performance heterogeneous controllers (the ones that provide different amounts of resource/actuation) for certain simple autonomous agent dynamics. Our efforts here show that similar tools—which are based on optimization machinery together with eigenvalue sensitivity and graph algebra notions—are applicable to an important family of decentralized design problems for both autonomous agent and internally coupled networks, including some constrained ones. Our results also highlight that very simple and structurally insightful design tools can be obtained for certain special classes of network topologies, such as the ones with non-negative weights.

The rest of the paper is organized as follows. In §2 we motivate the importance of decentralized design using several applications, including virus-spreading control and air traffic management. In §3, a canonical design problem that addresses these applications is formulated. Specifically, the problem of designing diagonal matrices D and/or K to minimize a cost that is a function of $D + KG$, where the square matrix G represents the network topology, is introduced. Section 4 introduces the analytical methods used to solve these design problems and summarizes the design; some details can be found in appendix A. Finally, examples are pursued in §5.

2. Controller synthesis problems in modern networks

Decentralized controller design is a critical need in myriad network applications. Here, we formulate network controller synthesis problems in some detail for two applications: (i) virus-spreading control and (ii) coordinated air traffic flow management (TFM). We also briefly describe synthesis problems in two other domains, namely autonomous vehicle control and iterative numerical solution of systems of linear equations. Each synthesis task is an example of the canonical problem pursued in this article. More extensive formulations of these four design tasks can be found in Roy & Saberi (2007), Wan & Roy (in press) and Wan *et al.* (submitted).

(a) Spatially heterogeneous virus-spreading control

The impacts of recent biological virus epidemics (SARS and foot-and-mouth disease) and computer viruses highlight the need for mathematically modelling virus spread and, in turn, developing controllers that reduce epidemic spread with limited resources. Epidemic control can be viewed as reducing the *basic reproduction ratio* R_0 (defined as the average number of secondary infections produced during an infected individual's infectious period, when the individual is introduced into a population where everyone is susceptible; Anderson & May 1991). For virus spreads in heterogeneous populations, R_0 is often computed as the dominant eigenvalue¹ of the *next generation operator (matrix)* of a *multi-group model* (see Diekmann & Heesterbeek 2000). Although spatial structure is believed to greatly impact epidemic spread (Riley *et al.* 2003), there is little work in the literature on designing controls (e.g. isolation and quarantine) that exploit the network's topological structure to reduce epidemic spread with limited resources. We seek to design limited resource allocations (controls) that optimally reduce R_0 by exploiting the topological structure.

In the interest of clarity, we pursue the design for a basic (susceptible–infected–susceptible) heterogeneous multi-group model, with the understanding that similar designs can be developed for more detailed models. Specifically, the multi-group model for spatial epidemic spread developed in Riley *et al.* (2003) can be generalized to admit spatially heterogeneous control. In this case, the next generation matrix is

$$A = \beta T \operatorname{diag}(T_i r_i N_i) \left(\operatorname{diag}(h_{ii}) + \operatorname{diag}(c_i) \begin{bmatrix} 0 & h_{21} & \dots & h_{n1} \\ h_{12} & 0 & \dots & h_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1n} & h_{2n} & \dots & 0 \end{bmatrix} \right) [\operatorname{diag}(N_i)]^{-1},$$

where β is the transmission coefficient (incorporating infectiousness and average contact rate); T is the nominal infectious period; N_i is the population in region i ; and each h_{ji} is a corrective term that accounts for the relative rate of inhomogeneous mixing between regions i and j . The region-specific scaling parameters T_i , r_i and c_i ($\in [0,1]$) are the controls that admit design.

¹ R_0 is a real number since an interaction network structure is non-negative and irreducible.

- (i) The parameter r_i scales the contact rate of individuals in region i , which decreases virus spread both locally and to spatial neighbours. Note that r_i can be reduced by isolation of closely connected and fairly isolated groups such as a school/college or by vaccination.
- (ii) The parameter c_i scales the contact rate of individuals from outside regions to a region i . The external contact rate factor c_i can be reduced by prohibition of travel from other regions to region i or, similarly, isolation of arriving travellers for longer than the incubation period.
- (iii) We allow the average duration of the infectious period in each region i to deviate from T by a factor of T_i , which can be reduced by shortening the time between symptom appearance and hospitalization in region i .

Control measures such as isolating people who may have contacted an infected individual (e.g. through isolation of a neighbourhood) reduce both T_i and r_i .

Our aim is to design $T_i r_i$ and/or c_i to minimize the basic reproduction ratio (i.e. the dominant eigenvalue of A) subject to the individual constraints on these parameters and subject to total resource constraints ($\sum T_i r_i$ or $\sum c_i$ larger than a lower bound). For instance, to design $T_i r_i$, we can form the following constrained optimization problem.

Problem 2.1. Design a diagonal matrix K , such that $\lambda_{\max}(KG)$ is minimized, where K is subject to the following constraints:

- (i) $\text{tr}(K) = \sum K_i \geq \Gamma$ and
- (ii) $0 \leq K_i \leq 1$ for all i .

Here, the *actuation matrix* K equals $\text{diag}(t_i r_i)$; the *topology matrix* G captures the remainder of the dynamics; the constraints on $T_i r_i$ yield that $0 \leq K_i \leq 1$ and that K_i in total exceed a *lower bound* Γ (since much resource is needed to make K_i small; see [Wan et al. \(submitted\)](#) for further details).

In the case that we design c_i , we can similarly formulate the following design problem.

Problem 2.2. Design a diagonal matrix K , such that $\lambda_{\max}(D + KG)$ is minimized, where K is subject to the following constraints:

- (i) $\text{tr}(K) \geq \Gamma$ and
- (ii) $0 \leq K_i \leq 1$ for all i .

Equivalently, we address the problem of reducing R_0 to a desirable value (e.g. the critical threshold $R_0 = 1$) using minimal amounts of resources. Such a systematic design of these inhomogeneous control parameters can provide guidance for effective epidemic control.

Automaton models that capture the evolution of *individuals'* infection states through interaction are also widely used in epidemic modelling, especially in characterizing computer viruses. We have also formulated the problem of designing optimal spatially heterogeneous controls in the context of an automaton model (in which probabilities that individuals are infected evolve). This decentralized design problem is similar to the one detailed above (see [Wan et al. \(submitted\)](#) for details and [Wang et al. \(2000\)](#) for background). We stress

here that the design problems introduced in this study are decentralized ones, in that the controls act on populations travelling to or from one region, or on individuals.

(b) Coordinated air TFM

One major complexity in air TFM is that flow restrictions (e.g. en route miles-in-trail restrictions, ground delay programs) only act locally, and yet must be coordinated to achieve network-wide objectives (Farley et al. 2001). Appropriately designing decentralized management strategies can permit significant reduction in traffic delays, while simultaneously reducing human-controller workload and reducing safety concerns.

The bulk of the literature on coordinated flow management either focuses on how strategies can be implemented in practice (Hoang et al. 2002) or pursues the design of coordinated TFM, assuming that aircraft flows are deterministic and further that individual aircraft can be scheduled at each restriction (Bayen et al. 2004). In fact, uncertainties (due to take-off time variations and weather) critically impact flows and also many flow restrictions simply enforce spacing between aircraft rather than permitting arbitrary scheduling. With these concerns in mind, we have developed a family of abstractions for restrictions acting on stochastic flows, each of which demonstrates the essential trade-off between downstream variability and upstream backlog/delay caused by a restriction (Wan & Roy in press). Here, we incorporate a highly abstracted algebraic model for a restriction into an Eulerian network flow model (which captures take-off/splitting/merging flows) to pose a network-wide flow management design problem. This formulation allows us to design restriction strengths and further yields graphical insights into the structure of good flow management designs.

Our model captures aircraft-count variabilities before and after n boundary restrictions (which we denote v_i and w_i , respectively), as well as the backlogs B_i caused by these boundary restrictions. We assume a simple linear dependence of the downstream variability and upstream backlog (in the steady state) on the strength of the restriction. Specifically, for boundaries within the airspace, we have $w_i = (1 - a_i)v_i$ and $B_i = \gamma a_i \lambda_i$, where $a_i \in [0, 1]$ is the restriction strength; γ is a scaling factor; and λ_i is the mean number of aircraft passing the boundary. For boundaries i corresponding to flows entering the airspace, we assume variabilities corresponding to Poisson process flows, and no backlog. We also model splitting/merging of flows between boundaries by relating each inflow variability to the outflow variabilities of neighbouring restrictions, specifically as $v_i = \sum_{j=1}^n w_j g_{ji}$, where $g_{ji} > 0$ implies that there is a flow from restrictions j to i . We note that some restriction strengths a_i can be designed, while others are fixed.

The design problem of interest is to set the parameters a_i to get desirable variability in downstream flows or regional counts, while maintaining small backlogs (so that total counts in upstream regions do not exceed thresholds and aircraft are not subject to long delays). Specifically, a natural performance measure is one that captures the total impact of the control on the aircraft counts in a couple of critical regions by combining the variabilities of flows in the regions with the backlogs in the regions caused by restrictions. Many such measures are well approximated as being linear or quadratic in the backlogs B_i and variabilities w_i .

The design problem posed above can straightforwardly be rewritten as the following linear algebraic design problem.

Problem 2.3. Design a diagonal matrix Z so as to minimize a cost measure that is quadratic in the entries of $I - Z$ or Z^{-1} and in \mathbf{w} (e.g. $(Z^{-1}\mathbf{1})^T(Z^{-1}\mathbf{1}) + \mathbf{w}^T\mathbf{w}$) subject to the following constraints:

- (i) the system of n equations $(I - Z\hat{G})\mathbf{w} = Z\hat{\lambda}$ is satisfied and
- (ii) each diagonal entry of Z is constrained to be in $[0,1]$.

Here, the topology matrix \hat{G} can be simply computed from the routing parameters g_{ji} ; the input vector $\hat{\lambda}$ depends on the boundary flow rates λ_i ; and the actuation matrix Z contains the designable restriction strengths a_i . We stress here that this design problem is also a decentralized one, in that each restriction impacts only local flows and yet a global objective is pursued.

(c) Other applications

(i) Controlling autonomous vehicle teams

Autonomous vehicle teams must move in formation for such diverse purposes, such as sweeping for mines, exploring underwater geological formations or studying the Martian landscape. Physical constraints, cost requirements and security concerns often necessitate that the vehicles only make decentralized measurements, i.e. each vehicle can only partially observe the entire network's state through some set of (in general non-absolute) position and velocity observations. The decentralized controller synthesis task then is to design the force inputs to each vehicle from its observations. Assuming a static (memoryless) feedback control paradigm, a high-velocity-gain controller can generally be used (see Roy & Saberi 2007). In this case, the high-performance design problem can be abstracted to the decentralized design problem of selecting diagonal matrix K to place the eigenvalues of KG at desirable locations, where G describes the agent's position observation topology. Analogous problems also arise in designing, for example, distributed estimation algorithms for sensor fusion.

(ii) Designing preconditioners for numerical analysis

Preconditioners are used to speed up convergence of linear iterative algorithms, such as those that can be used to solve the system of linear equation $Gx = b$ (Greenbaum 1997). Diagonal preconditioners are especially common because the extra computation required for their implementation is small. Essentially, a preconditioner's performance can be measured in terms of a *condition number* (typically either $|\lambda_{\max}(KG)|/|\lambda_{\min}(KG)|$ or $|\sigma_{\max}(KG)|/|\sigma_{\min}(KG)|$, where $\lambda(\cdot)$ and $\sigma(\cdot)$ are the eigenvalues and singular values, respectively, and K is the preconditioner). Thus, an optimal diagonal preconditioner is a diagonal matrix K that minimizes one of the condition numbers, subject to a constraint on the signs of the eigenvalues. Finding the optimum is a decentralized design problem, in that we are optimizing gains (weights) that have 'local' influence (change one row of G). The methodologies developed here can give insight into the structure of the optimal diagonal preconditioner.

3. The decentralized design problem

The controller design tasks introduced in §2 are all of the following form: a scalar cost $c(D+KG)$ must be minimized with respect to the diagonal actuation matrices D and K , where the topology matrix $G \in R^{n \times n}$ captures the interactions among components in a network, and the diagonal entries of D and K are subject to one or more constraints. We view this design problem as canonical but widely applicable, and hence seek methods for solving it.

Let us first elaborate on each aspect of this design problem and distinguish the problem from those previously addressed in the matrix analysis literature.

The topology matrix G represents the strengths and polarities of interactions among network components, such as traffic flow densities between boundaries in the airspace or infection-spread probabilities for networked computers. The topology matrix may in general be an arbitrary real, square matrix, but for many applications it is known to be more structured, for instance non-negative, symmetric and/or positive definite (Berman & Plemmons 1994). We note that the network's interaction topology can be illustrated using a graph with n nodes and with weighted and directed edges defined from the topology matrix.

The actuation matrices D and K directly or indirectly specify the control effort provided to each component in the network, for instance the quarantining/vaccination efforts provided to each district to prevent biological virus spread or the actuation of autonomous vehicles in a team. We stress here that D and K capture distributed control efforts in that they locally or partially change the structure of G . The additive matrix D describes entirely local actuations or topology changes in the network (e.g. the use of a computer virus detection and removal program). Meanwhile, the multiplicative actuation matrix K describes actuations or topology changes that proportionally impact all components connected to a particular component (e.g. a restriction in the air traffic system, which proportionally reduces flows to all downstream restrictions). The entries in the diagonal matrices D and K may be variously constrained; particularly common are instances where the entries are constrained to a simplex. Such a constraint results, for instance, when the total resource allocated for virus control is lower-bounded and the resource permitted in each region is bounded in an interval. We note that the special case where $D=0$, and hence the topology matrix has the form KG , is relevant in several application areas.

The *cost function* $c(\cdot)$ is application specific. However, the cost is very often associated with dynamics defined on the actuated network topology $D+KG$, as in all the examples in §2. In fact, the cost functions for the examples in §2 are all based on linear (or linearized) dynamics defined on the network. That is, each component in the network has dynamically varying states x_i associated with them (e.g. numbers of aircraft or probabilities of infection), such that $\mathbf{x}^T \triangleq [x_1, x_2, \dots, x_n]$ is governed by the differential equation $\dot{\mathbf{x}} = (D+KG)\mathbf{x} + \mathbf{w}$ or $\mathbf{x}[k+1] = (D+KG)\mathbf{x}[k] + \mathbf{w}$, where $\mathbf{w}(t)$ specifies external inputs to each component. Since these dynamics (whether representing virus spread or vehicle movement) are of fundamental interest for the applications, it is no surprise that the optimization cost is often defined based on the solution of the above linear differential/difference equations. Classically, the transient solutions of these equations are written as a sum of response terms each governed by an *eigenvalue* of the matrix $D+KG$, and hence the *stability* and *settling rate* of the dynamics

depend on the eigenvalue locations in the complex plane. This motivates the design of actuation matrices to optimize a dominant eigenvalue-based parameter (often the least negative among the eigenvalue's real parts in the continuous-time case, and the largest among the eigenvalue's magnitudes in the discrete-time case). Alternatively, the steady-state solutions to the above linear equations for a constant persistent input $\mathbf{w}(t) = \bar{\mathbf{w}}$ (given by $-(D + KG)^{-1}\bar{\mathbf{w}}$ and $(I - (D + KG))^{-1}\bar{\mathbf{w}}$, respectively, assuming stability) may define the cost.² It is worth noting that, even if dynamics are not of interest, the eigenvalues of $D + KG$ give much insight into the network's topological structure (see Chung 1994) and hence eigenvalue design is commensurate with topology design.

Connection to existing matrix algebra methods. This design problem unfortunately cannot be addressed by the existing matrix algebra techniques developed for complex networks. Spectral graph theory methods and methods for particular matrix classes largely focus on analysing a given topology (matrix), rather than seeking high-performance topologies among a class of possibilities (Berman & Plemmons 1994; Chung 1994). Meanwhile, the D -stability and diagonal Lyapunov stability tools give conditions under which all gains within a class achieve stability (essentially a robustness result), rather than identifying a single high-performance design.

4. Novel methodology for network controller design

Fundamentally, our control-theoretic method for solving the described network design problem is based on recognizing that the actuated topology matrix $D + KG$ has a special structure when the optimal actuation matrices $D = D^*$ and $K = K^*$ are chosen. That is, optimal or near-optimal actuations serve to alter the topology matrix to a particular form; our approach is to identify this form, which in turn permits us to compute the optimal actuation explicitly with a finite search. We have applied this three-step methodology to design both D and K for various cost functions, including those discussed in §2. In order to make the presentation accessible to a broad audience and also in the interest of space, here we include only an overview of the methodology and a brief description of the results. We kindly refer readers to Roy & Saberi (2007), Wan & Roy (in press) and Wan *et al.* (submitted) and appendix A for details. Specifically, we achieve the design through the following steps.

(a) Finding the structure of the optimally actuated topology

We apply standard constrained optimization techniques (i.e. Lagrange multiplier techniques; Bertsekas 1995) to obtain structural characteristics of the optimally actuated topology matrix $D^* + K^*G$. Specifically, by taking derivatives of the *Lagrangian* (which is formed from the cost function and constraints), we can obtain a set of equations that are necessarily satisfied by the optimization parameters (in our case, the diagonal entries of D and K). Here, we are very often concerned with costs based on dynamics and especially

²The careful reader will note that the asymptotic cost for the air traffic example is slightly different in form. The difference results because the inflow variabilities can be easily eliminated from the original expressions for the asymptotics in this case. Either form can serve as a starting point for the design.

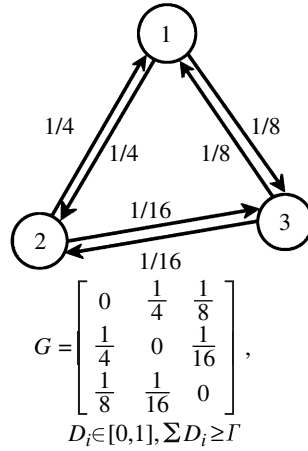


Figure 1. The structure of the optimally actuated topology is illustrated for the cost $\lambda_{\max}(D + G)$. Optimal designs: (i) $\Gamma = 1.5$: $D_1^* = 0.4167$, $D_2^* = 0.4792$, $D_3^* = 0.6042$, $\lambda_{\max}^* = 0.7917$ and $v_{\max}^* = [1 \ 1 \ 1]^T$. (Individual constraints not reached; all components participate equally.) (ii) $\Gamma = 2.9$: $D_1^* = 0.928$, $D_2^* = 0.972$, $D_3^* = 1$, $\lambda_{\max}^* = 1.2661$ and $v_{\max}^* = [1 \ 1 \ 0.7047]^T$. (Components 1 and 2 participate equally; component 3 at constraint.)

eigenvalues, and hence taking derivatives with respect to the optimization parameters involves invoking *eigenvalue sensitivity results* (Stewart 1974). In doing so, we obtain structural conditions on the *eigenvectors* of $D^* + K^*G$, i.e. when the optimizing actuations are used. This special structure implies to us that critical dynamics of the optimized network are excited by, and propagate in, certain spatial patterns. Identifying these common structural and dynamic characteristics of optimally actuated topologies is one of the key contributions of this research.

Let us list the structure of the optimally actuated topology, for several example costs and constraints.

- Consider designing D to minimize the *dominant eigenvalue* of $D + KG$, where each gain D_i is constrained to an interval and $\sum D_i$ is also constrained. For the optimizing actuation $D = D^*$, the following is true for each $i \in 1, \dots, n$: either D_i^* is at a limiting value or the i th *participation factor* (the product of the left and right eigenvectors' i th components³) of the dominant eigenvalue of $D^* + KG$ is equal to a fixed constant. That is, the optimizing actuation serves to equalize the participation factor of each network component (from actuation to response) in the dominant eigenvalue dynamics, to the extent permitted by the constraints (figure 1). We refer the readers to theorem A.1 in appendix A for a formal statement of this result and also the proof. Briefly, the result is developed as follows: for an optimizing actuation, from Lagrange multiplier techniques, we obtain that either the sensitivity of the dominant eigenvalue to each parameter is 0 or the parameter hits the constraint boundary. Working from the fact that the dominant eigenvalue's sensitivity to

³ Participation factors were originally introduced in the study of electric power networks, to codify how the dynamics of a particular mode (eigenvalue) of a system are impacted by each network component (see Perez-Arriaga et al. 1990).

each parameter (that is not at a boundary) is 0, we obtain from the left and right eigenvector equations (with a little algebra) that the corresponding participation factors are identical.

- For optimal diagonal preconditioners K^* , the i th participation factors of the maximum and minimum eigenvalues of K^*G are equal in absolute value (i.e. they can differ only in sign). The development of this result is analogous to the one for the dominant eigenvalue of $D+KG$ (see theorem 3 in Roy & Saberi (2007) for details).
- For the asymptotic cost $\mathbf{1}^T(I-(K+G))^{-1}\bar{\mathbf{w}}$ subject to a simplex constraint, the product of the asymptotic state \bar{x}_i with the sum of the i th column of $(I-(K+G))^{-1}$ is equal for all i . The methodology for deriving this result is again very similar to that for the above costs, with only the exception that the sensitivity of the quadratic cost rather than of an eigenvalue is used.

(b) Computing the optimal actuations

By invoking these structural characteristics of the optimally actuated topology, it turns out that we can limit ourselves to solving a finite (albeit, in general, large) set of systems of equations to find the optimum actuations D^* and K^* . Thus, we can give explicit formulae for the optimizing actuations or develop finite search algorithms for finding them. We note that these reformulations often require the use of basic eigenanalysis or algebraic graph theory notions such as the Courant–Fischer theorem (Chung 1994). For instance, for the optimization of $\lambda_{\max}(D+G)$ with respect to D , the structural result described above allows us to find D^* by choosing each D_i as 0, 1 or $0 < D_i < 1$. By placing sets of D_i at 0 and 1, respectively, and then solving for the remaining gains based on the structural result, we find that the optimal actuations can be computed by solving at most 3^n systems of equations (see theorem A.4 in appendix A).

(c) Simplifying computation for special classes of topologies

For special classes of topology matrices (e.g. non-negative, diagonally symmetrizable and/or positive definite; Berman & Plemmons 1994; Gallager 1996), the dimension of the computation of D^* or K^* can often be reduced even further. These simplifications result because the special topology guarantees convexity of the optimization (so local optimality yields global optimality) and/or because the eigenvectors of $D+KG$ are further structured in these cases. The additional structure also often permits simple interpretation of the optimizing actuations.

For instance, again, consider optimizing the dominant eigenvalue of $D+KG$ with respect to D , subject to the simplex constraints. When G is non-negative and diagonally symmetrizable (which are reasonable assumptions in describing, e.g. virus spread), the iterative algorithm for finding D^* in many cases requires at most n solutions of systems of equations rather than 3^n . Specifically, the optimal actuation can be found by first finding the optimum when the constraints on individual D_i are relaxed. Then, if too much actuation is required for some i , these D_i can simply be set to their limiting value and the optimum recomputed (with the process repeated if other gains exceed their bounds; see theorem A.6 in appendix A). An eigenvector majorization result allows us to justify that such a simple algorithm can find the optimum. Recently, we have been able to obtain

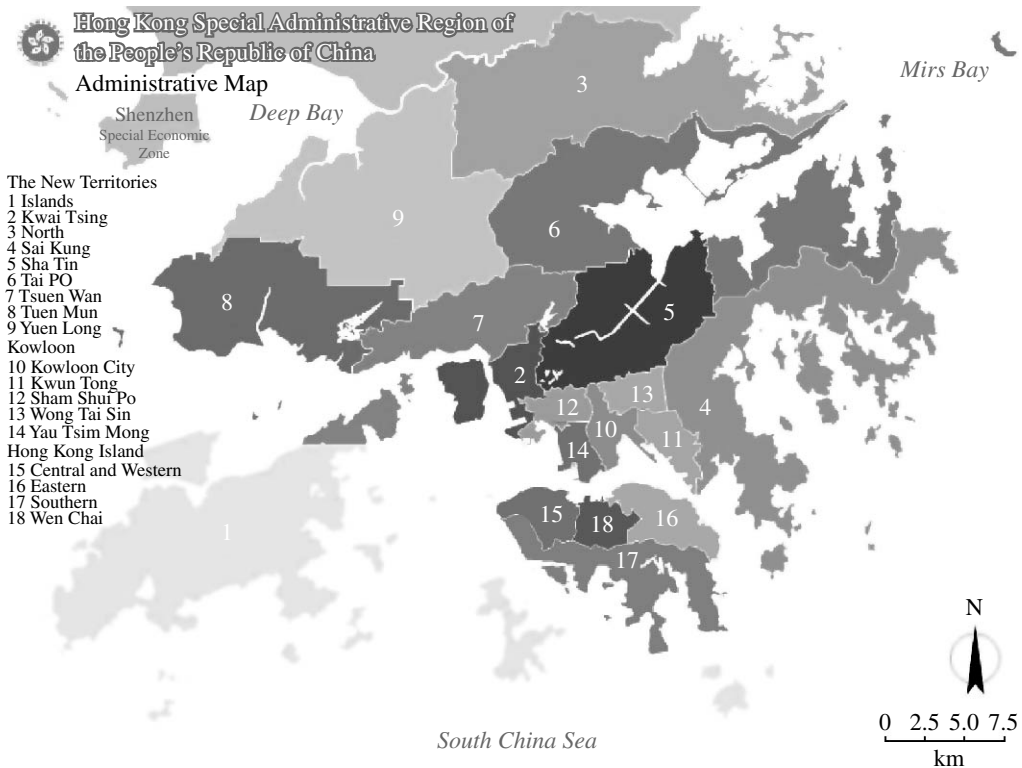


Figure 2. The map of Hong Kong's 18 districts (obtained from http://en.wikipedia.org/wiki/Hong_Kong#Administrative_divisions). We use the model parameters given in Riley (2003): $h_{ii}=1$, $h_{ij}=0.57$ when districts i, j are neighbours; $h_{ij}=0.02$ when i, j are not adjacent; $\beta=0.062$; and $T=10.6$.

similar results on majorization of eigenvector components for arbitrary irreducible non-negative matrices (not just diagonally symmetrizable ones), which permit the extension of this algorithm to such topologies (Roy *et al.* submitted).

In terms of simplifying interpretations, for positive and symmetric topologies, one can see that the requirement of equalized participation factors corresponds to seeking equalization of row sums, or individual components' impacts, as best as is permitted by the constraints (lemma A.5). Furthermore, the proof of theorem A.6 also clarifies that, for non-negative topologies, a network component that cannot fully participate due to its resource limits is supported through extra resource allocation at its neighbours. This interpretation is, for instance, valuable in virus-spreading applications.

5. Examples

(a) Control of the Hong Kong SARS epidemic

The article by Riley *et al.* (2003) developed a multi-group model for the propagation of SARS in Hong Kong's 18 districts (figure 2) and identified the homogeneous control needed to reduce the basic reproduction ratio R_0 to 1. Our approach allows us to find the optimal heterogeneous control that uses the same

Table 1. Resource allocation needed to reduce R_0 to 1. (The optimal resource allocation for each region is shown, and the total resource allocation is compared with the homogeneous strategy. The entries in italics indicate that the corresponding regions receive the most resources.)

	$h_{ii}=1$	$h_{ii}=1.5$	$h_{ii}=3$
district 1	0.0904	0.3778	0.6969
district 2	0.0904	0.4632	0.7475
district 3	0.4184	0.5662	0.7454
district 4	<i>1</i>	<i>1</i>	0.8081
district 5	<i>1</i>	<i>0.984</i>	<i>0.9085</i>
district 6	0.4184	0.5758	0.8295
district 7	<i>1</i>	<i>1</i>	<i>0.9166</i>
district 8	0.0904	0.4074	0.7292
district 9	<i>1</i>	<i>0.9211</i>	0.8266
district 10	<i>1</i>	<i>1</i>	0.8581
district 11	0.4184	0.5447	0.7806
district 12	<i>1</i>	0.7882	0.8202
district 13	0.4184	0.5543	0.8015
district 14	0.0904	0.4572	0.7413
district 15	0.0904	0.5216	0.7525
district 16	0.0904	0.5216	0.7525
district 17	<i>1</i>	0.8081	0.8092
district 18	<i>1</i>	0.8081	0.8092
total for heterogeneous control	10.22	12.30	14.33
total for homogeneous control	12.99	13.60	14.78

total resource amount (see also Wan *et al.* submitted); this control reduces the basic reproduction ratio to 0.64. Equivalently, it is easily shown that $R_0=1$ can be achieved even when the total control resource is reduced to 79% of the one with equal allocation (table 1).

These studies show that an epidemic's spread can be stopped more quickly with the same total control resources through heterogeneous allocation. This intelligent allocation takes advantage of the spatial structure of the population by placing more control resources in the highly connected parts of the network, such as districts 5 and 7 in Hong Kong, and fewer resources in isolated parts such as district 1 (figure 2; table 1). In this way, the limited control resources are best able to reduce the rate at which the epidemic diminishes. Such a control would significantly reduce the impact on people's daily lives in some districts (which have less control resources allocated) and overall, while still stopping the virus spread quickly.

For illustration, we also consider how the heterogeneous resource allocation changes when the local mixing rate h_{ii} is increased (compared with the mixing rate of neighbouring districts h_{ij}). As expected, increasing the local mixing rate makes the resource allocation more homogeneous (table 1).

(b) TFM design

We also use our methodology to design en route restrictions for air TFM (see illustration of the restriction model in figure 3). Specifically, we consider restriction design for several congested regions in an air traffic network, with merging and

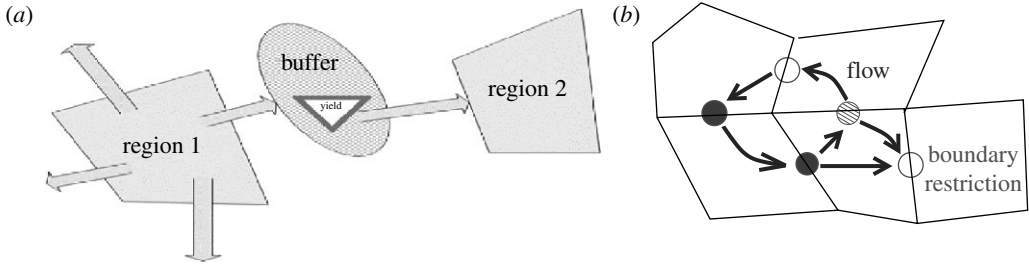


Figure 3. (a) The restriction model is illustrated. A restriction holds aircraft at boundaries. Here, an algebraic model for restrictions is considered. (b) Traffic flows between boundaries in congested regions are shown. Also, the optimal restriction topology is given as follows: dark shading, strong restriction; light shading, weak restriction.

splitting flows (figure 3). Flow management is required to reduce the variability in regional aircraft counts (so as to prevent capacity violations), but without causing excessive backlog and delay. As discussed in §2, the flow management problem can be abstracted to the decentralized design problem studied here. We can thus use the design methodology to optimally choose restriction strengths in this example, as shown in figure 3. In agreement with detailed modelling-based and operational studies (Wan & Roy in press), our approach shows that restrictions are most effective when placed upstream on flows that impact multiple congested regions. The additional value provided by our approach is the ability to design restriction strategies for arbitrary flow topologies and the possibility of optimizing/improving these strategies.

This work was partially supported by the National Science Foundation and the National Aeronautics and Space Administration.

Appendix A

We have applied the design methodology introduced here to several specific problems with various costs and constraints (see Roy & Saberi 2007; Wan & Roy in press; Wan et al. submitted). Here, for reader convenience, we illustrate the methodology with proof for one canonical example, namely the design of diagonal D to minimize the dominant eigenvalue of $D+G$ (where we have renamed KG as G for convenience, since K is assumed fixed in this case) subject to the constraints that $0 \leq D_i \leq L$ and $\text{tr}(D)$ is lower-bounded (see also Wan et al. submitted). This particular form of the canonical design problem is applicable to virus-spreading control in computer networks, among other applications.

For convenience, we refer to the optimum D as D^* , the dominant eigenvalue of D^*+G as λ_{\max}^* and the corresponding left and right eigenvectors as w_{\max}^* and v_{\max}^* . We begin with a structural result.

Theorem A.1. Consider a matrix $D+G$, where D is diagonal and G is an $n \times n$ matrix, and consider $D=D^*$ that minimizes the dominant eigenvalue of $D+G$ subject to the constraints (i) $\sum D_i \geq \Gamma$ and (ii) $D_i \in [0, L]$, and assuming the

dominant eigenvalue of $D+G$ to be real and non-repeated,⁴ the optimizing D^* and corresponding eigenvalue/eigenvectors λ_{\max}^* , w_{\max}^* and v_{\max}^* (appropriately normalized) satisfy one of the following conditions. For either $\sum D_i^* = \Gamma$ and for each i we have either $0 < D_i^* < L$ and $w_{\max,i}^* v_{\max,i}^* = 1$ or $D_i^* = L$ or $D_i^* = 0$, or for $\sum D_i^* < \Gamma$ and for each i we either have $0 < D_i^* < L$ and $w_{\max,i}^* v_{\max,i}^* = 0$ or $D_i^* = L$ or $D_i^* = 0$.

Proof. This proof follows from standard theorems on eigenvalue sensitivity (Stewart 1974), as well as theorems on constrained optimization using Lagrange multipliers (Bertsekas 1995). Let us denote $D_i = d_i^2$, since we require $D_i \geq 0$ for all i . The procedure for finding the optimum D^* under constraint is to form the Lagrangian $L = \lambda_{\max}(D+G) + \sum a_i(d_i^2 + m_i^2 - L) - C(\sum d_i^2 - n^2 - \Gamma)$ and set the derivatives of it with respect to all variables (namely d_i , a_i , m_i , C and n_i) to 0. (Here, m_i and n are slack variables to transform inequality constraints to equality constraints.) This procedure leads to the following equations:

$$\left. \begin{aligned} d_i^*(w_{\max,i}^* v_{\max,i}^* + a_i^* - C^*) &= 0, \\ d_i^{*2} + m_i^{*2} &= L, \\ m_i^* a_i^* &= 0, \\ n^* C^* &= 0 \end{aligned} \right\} \quad (\text{A } 1)$$

and

$$\Gamma + n^{*2} = \sum d_i^{*2}.$$

Note that the first equation above follows from the eigenvalue sensitivity formula. The two cases in the theorem thus follow automatically from consideration of the variables n^* and C^* , one of which must be 0. Specifically, the case where $\sum D_i^* = \Gamma$ follows from setting n^* to 0, while the case where $\sum D_i^* > \Gamma$ follows from setting C^* to 0. ■

Theorem A.1 states that the optimum D^* can be either at or inside the constraint boundaries. When D^* is at the boundary $\sum D_i^* = \Gamma$, each D_i^* falls into one of the three categories: (i) $D_i^* = L$, (ii) $D_i^* = 0$ and (iii) $w_{\max,i}^* v_{\max,i}^* = 1$. When D^* is not at the boundary $\sum D_i^* = \Gamma$, each D_i^* again is at 0 or L , or $w_{\max,i}^* v_{\max,i}^* = 0$. We note that, for any condition in one of the above forms, the number of equations and variables is equal, and so we can get a potential optimal solution from these equations. However, the number of potential optimal solutions (the number of solutions that satisfies one set of equations of this type) grows exponentially with the dimension of the matrix G , and so the calculation will be very complicated for even moderate-sized G . In the rest of this section, we will show that when G is specially structured, e.g. non-negative or symmetric, we can develop more explicit (and hence easier-to-evaluate and interpret) expressions for the optimal solution.

In the following theorem A.2, we show that for a very broad class of topology matrices G , the optimizing D is one that uses maximum total resources, i.e. one for which $\sum_i D_i = \Gamma$.

⁴The theorem can be easily generalized to the case that $D+G$ has real and simple dominant eigenvalues.

Theorem A.2. Consider $D+G$, where D is diagonal, G is an $n \times n$ matrix, and the largest eigenvalue of $D+G$ is real and non-repeated for all D , such that (i) $\sum D_i \geq \Gamma$ and (ii) $D_i \in [0, L]$. The matrix D^* that minimizes the dominant eigenvalue of $D+G$ subject to these constraints satisfies $\sum D_i^* = \Gamma$, if the left and right eigenvectors of $D+G$ corresponding to the dominant eigenvalue have the same sign patterns for all D . Classes of matrices satisfying this condition include (i) irreducible non-negative matrices and (ii) diagonally symmetrizable matrices (matrices for which there exists a diagonal Q , such that $Q^{-1}GQ$ is symmetric).

Proof. The condition that the left and right eigenvectors of the dominant eigenvalue have the same sign pattern implies the relationship that, for all i , $w_{\max,i} v_{\max,i} > 0$. Thus, according to the eigenvalue sensitivity theorem, the dominant eigenvalue decreases monotonically with the decrease of D_i for all i , since $(\partial \lambda_{\max}(D+G))/\partial D_i = w_{\max,i} v_{\max,i}$ is positive. Therefore, the optimum D^* is on the boundary $\sum D_i^* = \Gamma$. From the Perron–Frobenius theorem, the dominant eigenvalue of any non-negative and irreducible matrix is real and non-repeated, and the left and right eigenvectors associated with the dominant eigenvalue are positive (Gallager 1996), and hence have the same sign pattern. For a diagonally symmetrizable G , it is easy to check through a similarity transform that the eigenvalues are real and the left and right eigenvectors associated with any eigenvalue are identical related by a positive diagonal scaling, and hence have the same sign pattern. Hence, the theorem is proved. ■

Theorem A.2 guarantees that the optimum D^* is located on the boundary $\sum D_i^* = \Gamma$, whenever G is an irreducible and non-negative square matrix, and hence simplifies the search for D^* when G has this special structure. This simplification is relevant to our applications, because for both the multi-group and automaton models, G is non-negative and (for typical interaction topologies) irreducible. In the illustrating example (figure 1), G is irreducible and non-negative, so we expect that the optimum D^* satisfies $\sum D^* = \Gamma$. In this case, D^* can be found by searching only through the first set of possibilities in theorem A.1. This search is formulated in theorem A.4 for some matrices of this sort. Before that, in lemma A.3, we characterize the pattern of the eigenvector associated with the dominant eigenvalue under constraints $D_i \in [0, L]$ and $\sum D_i \geq \Gamma$.

Lemma A.3. Consider the matrix $D+G$, where D is diagonal and G is an $n \times n$ irreducible non-negative symmetric matrix. A matrix $D=D^*$ minimizes the dominant eigenvalue of $D+G$ subject to the constraints $\sum D_i \geq \Gamma$ and $D_i \in [0, L]$ if and only if the eigenvector v associated with the dominant eigenvalue of D^*+G has the pattern $v_{\max,i}(\forall i, \text{ s.t. } D_i=L) < v_{\max,i}(\forall i, \text{ s.t. } 0 < D_i < L) < v_{\max,i}(\forall i, \text{ s.t. } D_i=0)$, and $v_{\max,i}(\forall i, \text{ s.t. } 0 < D_i < L)$ are identical.

Proof. First let us show necessity. Suppose v_{\max}^* is the eigenvector associated with the dominant eigenvalue of D^*+G . The conclusion that $v_{\max,i}^*(\forall i, \text{ s.t. } 0 < D_i < L)$ are the same directly follows from theorem A.1 and the symmetry and non-negativity of G . More specifically, when G is a symmetric matrix, so is $D+G$, and thus $v_{\max,i}^* = w_{\max,i}^*$ for all i . Also, from theorem A.1, we know that the eigenvectors v_{\max}^* and w_{\max}^* of D^*+G satisfy $v_{\max,i}^* = w_{\max,i}^*$ and $v_{\max,i}^* w_{\max,i}^* = 1$ for each i , such that $0 < D_i^* < L$. Finally, from positivity of G , we see that v_{\max} has only positive entries. Combining, we recover that $v_{\max,i}^*(\forall i, \text{ s.t. } 0 < D_i < L)$ are

identical. Now we need to show that $v_{\max,i}^*(\forall i, \text{ s.t. } D_i=L) < v_{\max,i}^*(\forall i, \text{ s.t. } 0 < D_i < L) < v_{\max,i}^*(\forall i, \text{ s.t. } D_i=0)$. Since D^* achieves the minimum dominant eigenvalue, decreasing D_i^* for i , such that $D_i^* = L$ (making them less than L), or increasing D_i^* for i , such that $D_i^* = 0$ (making them larger than 0), while maintaining $\sum D_i = \Gamma$ should increase λ_{\max} . Eigenvalue sensitivity naturally leads to the inequality of eigenvector components, since the derivative of λ_{\max} with respect to D_i equals $v_{\max,i}^2$.

For the sufficient condition, we know that if $v_{\max,i}(\forall i, \text{ s.t. } D_i=L) < v_{\max,i}(\forall i, \text{ s.t. } 0 < D_i < L) < v_{\max,i}(\forall i, \text{ s.t. } D_i=0)$ and $v_{\max,i}(\forall i, \text{ s.t. } 0 < D_i < L)$ are the same, the corresponding D achieves a local minimum (from above). It follows from convexity (which can be proved easily using, e.g. the Courant–Fischer theorem) that the local minimum is in fact global. ■

Lemma A.3 presents the pattern of eigenvectors associated with the minimized dominant eigenvalue of $D+G$. This allows us to check whether a solution D is optimum by simply evaluating the dominant eigenvectors of $D+G$.

Theorem A.4. *Consider a topology matrix G that is non-negative, irreducible and diagonally symmetrizable. The matrix D^* that minimizes the dominant eigenvalue of $D+G$ can be found using the following algorithm.*

- (i) Find a diagonal matrix Q , such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as \hat{G} .
- (ii) Choose some D_i as 0 and some D_i as 1, and rearrange the rows and columns of $D + \hat{G}$ in such a way that those D_i fixed at 0 and 1 are in the lower right corner. The permuted matrix can be decomposed as

$$\begin{bmatrix} \hat{G}_{11} + D_A & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} + D_B \end{bmatrix},$$

where the matrix D_A is unknown and the matrix D_B has entries fixed at 0 or 1.

- (iii) Solve the equation $-\mathbf{1}^T \hat{G}_{11} \mathbf{1} - \mathbf{1}^T \hat{G}_{12} (\lambda \mathbf{I} - \hat{G}_{22} - D_B)^{-1} \hat{G}_{21} \mathbf{1} + \mathbf{1}^T \lambda \mathbf{1} = \Gamma - \text{tr}(D_B)$ for λ . This can be done through a simple numerical procedure.
- (iv) Calculate D_A using $D_A = \text{diagonalize}(-\hat{G}_{11} \mathbf{1} - \hat{G}_{12} (\lambda \mathbf{I} - \hat{G}_{22} - D_B)^{-1} \hat{G}_{21} \mathbf{1} + \lambda \mathbf{1})$. If $0 \leq D_A \leq LI$, and the pattern of eigenvector associated with the dominant eigenvalue of $D + \hat{G}$ follows lemma A.3, the optimum D is $\text{diag}(D_A, D_B)$. Otherwise, go to step (ii) until a solution is achieved.

Proof. We know from similarity that the eigenvalues of $D+G$ are the same as those of $D + \hat{G}$, where $\hat{G} = Q^{-1}GQ$ and Q is the positive diagonal matrix, such that \hat{G} is symmetric. Hence, we can, without loss of generality, consider designing D^* to minimize $D^* + \hat{G}$. Since all v_{\max} (of $D^* + G$) whose corresponding D_i^* are not fixed at 0 or 1 are equal (let us normalize them to 1), we know the optimum D satisfies

$$\begin{bmatrix} \hat{G}_{11} + D_A & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} + D_B \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \mathbf{v} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{1} \\ \mathbf{v} \end{bmatrix},$$

where D_B contains the entries in D which are 0 and 1. D_A must be found, and \hat{G}_{11} ,

\widehat{G}_{12} , \widehat{G}_{21} and \widehat{G}_{22} are submatrices of \widehat{G} (upon appropriate permutation). A little bit of algebra leads to the solution of λ , D and v . If the pattern of the eigenvector associated with the dominant eigenvalue of $D + \widehat{G}$ follows lemma A.3, and the D matrix is within the constraints $0 \leq D \leq LI$, we have found a global optimum solution according to lemma A.3. ■

Although we have presented an algorithm for the diagonally symmetrizable G , a slightly more complicated algorithm exists for all G that are non-negative matrices; we omit this algorithm here in the interest of space. We also note that the above algorithm may be computationally intensive in that the steps may have to be repeated up to 3^n times to find the optimum. For some topology matrices G , the calculation of D^* can further be simplified. We describe the procedure to calculate D^* under these circumstances in theorem A.6. As a preliminary step, let us first explicitly compute the optimal D when the constraints on individual D_i are relaxed.

Lemma A.5. *Consider the matrix $D + G$, where D is diagonal and G is an $n \times n$ symmetric irreducible and non-negative matrix. \bar{D} that minimizes the dominant eigenvalue of $D + G$ subject only to the constraint $\sum D_i \geq \Gamma$ can be found as follows: first, find $\bar{\lambda}_{\max} = (1/n)(\Gamma + \sum_i \sum_j G_{ij})$ and then find $\bar{D}_i = \bar{\lambda}_{\max} - \sum_j (G_{ij})$.*

Proof. First, we note that all entries of \bar{v}_{\max} are identical, based on lemma A.3 and the fact that here only the constraint $\sum D_i \geq \Gamma$ is enforced. Thus, the optimizing \bar{D} and eigenvalue satisfy $(\bar{D} + G)\mathbf{1} = \bar{\lambda}_{\max}\mathbf{1}$. Therefore, the \bar{D}_i values make the row sums of $D + G$ equal. \bar{D}_i also satisfies $\sum \bar{D}_i = \Gamma$, since G and thus $\bar{D} + G$ are symmetric. A little bit of algebra leads to the expressions for $\bar{\lambda}_{\max}$ and \bar{D}_i . ■

Lemma A.5 states that, without the constraints that $D_i \in [0, L]$, the optimum D (denoted \bar{D}) is the one that equalizes the row sum of $D + G$, i.e. a resource is allocated to each part of the network so as to make all their impacts identical. However, when the individual D_i is constrained, sometimes the optimum \bar{D} cannot be reached. Building on lemma A.5, theorem A.6 states an easy way to find the optimum D under several circumstances.

Theorem A.6. *Consider a matrix $D + G$, where D is diagonal and G is an $n \times n$ non-negative, irreducible and diagonally symmetrizable matrix. We can find $D = D^*$ that minimizes the dominant eigenvalue of $D + G$ subject to the constraints (i) $\sum D_i \geq \Gamma$ and (ii) $D_i \in [0, L]$ using the following algorithm.*

- (i) Find a diagonal matrix Q , such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as \widehat{G} .
- (ii) Calculate $\bar{\lambda}$ and \bar{D}_i following lemma A.5 and check if $0 \leq \bar{D}_i \leq L$ for all i . Denote the set of indices i such that $\bar{D}_i > L$ as \mathcal{L}^+ , and the set such that $\bar{D}_i < 0$ as \mathcal{L}^- . If $\mathcal{L}^+ = \emptyset$ and $\mathcal{L}^- = \emptyset$ (i.e. both sets are empty), we have $\lambda_{\max}^* = \bar{\lambda}$ and $D_i^* = \bar{D}_i$.
- (iii) If $\mathcal{L}^+ \neq \emptyset$ and $\mathcal{L}^- = \emptyset$, we set $D_i = L$, for $i \in \mathcal{L}^+$. That is, D_i can be calculated by first applying the following cases: $D_i = L$, for $i \in \mathcal{L}^+$; $v_{1_i} = 1$, for $i \notin \mathcal{L}^+$; $\sum_i D_i = \Gamma$; and solving the above equations for D_i as in theorem A.4. If $D_i \leq L \forall i$, we have $D^* = D$. Otherwise, we must recursively reduce those D_i , such that $D_i > L$, to L and recompute D until all $D_i \leq L$.
- (iv) If $\mathcal{L}^- \neq \emptyset$ and $\mathcal{L}^+ = \emptyset$, we set $D_i = 0$, for $i \in \mathcal{L}^-$. That is, D_i can be

calculated by first applying the following cases: $D_i = 0$, for $i \in \mathcal{L}^-$; $v_{1_i} = 1$, for $i \notin \mathcal{L}^-$; $\sum_i D_i = \Gamma$; and solving the above equations for D_i as in theorem A.4. If $D_i \geq 0 \forall i$, we have $D^* = D$. Otherwise, we must recursively increase those D_i , such that $D_i < 0$, to 0 and recompute D_i until all $D_i \geq 0$.

Proof. When $\mathcal{L}^+ = \emptyset$ and $\mathcal{L}^- = \emptyset$, obviously $D^* = \bar{D}$ is the optimum solution, as shown in lemma A.5.

When $\mathcal{L}^- = \emptyset$ and $\mathcal{L}^+ \neq \emptyset$, the optimal solution has the special feature that $D_i^* = L$, for $i \in \mathcal{L}^+$. We show this in the following.

Let us say the unconstrained optimum (\bar{D}) yields m elements in \mathcal{L}^+ (m of the \bar{D}_i are greater than 1) and let us consider the optimal solution when these \bar{D}_i are constrained to L (while the other entries are left unconstrained for now). We need to prove that the eigenvector components associated with these m elements are smaller than the eigenvector components associated with all the other elements (e.g. $\hat{v}_{\max, i \in \mathcal{L}^+} < \hat{v}_{\max, i \notin \mathcal{L}^+}$), and hence prove that this optimum is in fact the global one according to lemma A.3. We show this using induction.

Basis. Let us use $Z = \bar{D} + \bar{G}$ for the matrix that produces the unconstrained global optimal, call its dominant eigenvalue $\bar{\lambda}$, and assume (wlog) that the first m entries of D are too big (e.g. larger than L). Let us first consider decreasing the gain in entry 1 to unity while optimizing the other gains (right now, we do not apply constraints to these gains). Here, we denote D_i as gain in entry i . Then, the matrix of interest changes from Z to

$$Z + \begin{bmatrix} -\alpha & & \\ & \Delta D_2 & \\ & & \Delta D_n \end{bmatrix},$$

where $\alpha > 0$ and $\sum_{i=2}^n \Delta D_i = \alpha$. We would like to prove two statements: first, the eigenvector component corresponding to entry 1 is less than the components associated with all the other entries (e.g. $v_{\max, 1} < v_{\max, i \neq 1}$) and, second, the gain corresponding to entries from 2 to m is still at least L .

Recalling theorem A.1, we know that all $v_{\max, i \neq 1}$ are the same, say c_1 . For convenience, let us denote $v_{\max, 1}$ as v_1 . Hence, the eigenvector associated with the dominant eigenvalue (denoted by $\hat{\lambda}$) for the new matrix is $\hat{v} = [v_1 \ c_1 \ \dots \ c_1]$. The Courant–Fischer theorem (Chung 1994) states that $\bar{\lambda} = \max_v v^T Z v$ and $\hat{\lambda} = \max_v v^T (Z + \Delta) v$. The eigenvalue $\hat{\lambda}$ can be written as $\hat{\lambda} = \max_{v_1, c_1} (\hat{v}^T Z \hat{v} - v_1^2 \alpha + c_1^2 \alpha)$. When we decrease the gain in entry 1, we know that the maximum eigenvalue of the new matrix, say $\hat{\lambda}$, is larger than $\bar{\lambda}$. Moreover, we also know that $\hat{v}^T Z \hat{v} < \bar{\lambda}$, since \hat{v} is not the eigenvector for $\bar{\lambda}$. Thus, from the expression for $\hat{\lambda}$, the first entry in the eigenvector for the new matrix is smaller than the rest of the entries (i.e. $v_1 < c_1$). Thus, the first statement is proved.

For the second statement, we prove it as follows. Since $v_1 < c_1$, the eigenvector associated with the optimal gain changes from $\mathbf{1}_n$ to

$$\mathbf{1}_n + \begin{bmatrix} v \\ c \mathbf{1}_{n-1} \end{bmatrix},$$

where we know that $v < 0 < c$ (since only one gain has been moved so far). Furthermore, we know that the maximum eigenvalue of the new matrix, say $\hat{\lambda}$, is larger than $\bar{\lambda}$. Plugging into the eigenvector equation and doing some algebra, we

finally obtain the following equation:

$$(\hat{\lambda} - \lambda)\mathbf{1}_{n-1} + c(\hat{\lambda}\mathbf{1}_{n-1} - Z_{2:n,2:n}\mathbf{1} - Z_{2:n,1}(v/c)) = (1+c) \begin{bmatrix} \Delta D_2 \\ \vdots \\ \Delta D_n \end{bmatrix}.$$

However, since $\hat{\lambda} > \lambda$, $v < 0 < c$ and Z is a non-negative matrix, we recover that all entries in the expression on the l.h.s. are positive, and hence the changes in gains

$$\begin{bmatrix} \Delta D_2 \\ \vdots \\ \Delta D_n \end{bmatrix}$$

must all be positive and so each other gain strictly increases. Thus, we see that the gain corresponding to entries from 2 to m are still at least L .

Induction. Suppose that we bring any l of the $D_i > L$ to L ($l < m$). Let us assume first that the eigenvector components corresponding to the l entries are less than the components associated with all the other entries (e.g. $v_{\max, 1 \dots l} < v_{\max, i > l}$), and, second, that the gain corresponding to entries other than these l ones are still at least L . Let us show that after we bring another $(l+1)$ st D_i to L , we still have the appropriate eigenvector component majorization and condition on the gains.

Let us consider bringing $l+1$ of \bar{D}_i from the unconstrained optimum, such that $i \in \mathcal{L}^+$, to L . We can do this using two steps. The first step is to move all but one of the offending gains to L (l gains) and the second step is to bring the last gain to L . First note that this is possible, since after the first step, all other gains remain greater than L by assumption. Without loss of generality, for notational convenience, let us assume that we first bring the first l of \bar{D}_i , such that $i \in \mathcal{L}^+$, to L and then move \bar{D}_{l+1} . Denote the dominant eigenvalue after the first step as λ_1 and the one that after bringing the last gain to L as λ_2 . Again, applying the Courant–Fischer theorem, $\lambda_1 = \max_v v^T (D + \hat{G}) v$ and $\lambda_2 = \max_v v^T (D + \hat{G} + \Delta) v$, where the diagonal matrix Δ , corresponding to changing \bar{D}_{l+1} to L , has entries as follows: $\Delta_{l+1, l+1} = -b$ ($b > 0$); $\Delta_{i, i} = 0$, for $i \in [1 \dots l]$; and $\Delta_i = \Delta D_i$, for $i \in [l+2 \dots n]$. We also have $\sum_{i=l+2}^n \Delta D_i = b$. With a similar argument to that given in the basis argument, we can show that the $(l+1)$ st eigenvector component is less than the (identical) components after position $l+1$. Repeating this argument with each \bar{D}_i out of the $l+1$ possibilities set to L last, we can reach the conclusion that the eigenvector components corresponding to all the $l+1$ entries are less than the remaining common entries.

The proof that the remaining gains D_i increase (and hence that they remain larger than L if they were originally larger than L without the constraints) after bringing these $l+1$ gains D_i to L is based on the knowledge that the eigenvector components corresponding to all the $l+1$ entries are less than the remaining (identical) entries. This can be proved formally in a very similar fashion to the case where a single gain is moved, which we have addressed in the basis step of the induction. Thus, the details are omitted.

In case some other gains exceed L in the process, these can be reduced in the same fashion.

For the case that $\mathcal{L}^- \neq \phi$ and $\mathcal{L}^+ = \phi$, the proof is analogous to the case $\mathcal{L}^- = \phi$ and $\mathcal{L}^+ \neq \phi$, which we have proved here, and hence it is omitted. ■

Theorem A.6 provides an easy way to calculate the diagonal matrix D that minimizes the dominant eigenvalue of $D+G$ for a diagonally symmetrizable and non-negative G . We can first calculate the optimum \bar{D} without the individual constraints on D_i , i.e. if every \bar{D}_i satisfies its constraint, we have found the optimum. Otherwise, if the \bar{D}_i that violate their constraints are either all larger than L or all less than 0, the actual optimal D_i for positions where constraints are violated are equal to boundary values. This allows us to quickly locate the D_i values at the boundary rather than to try all combinations of k at the boundary to find the optimal solution. In fact, at most n cases (rather than 3^n) need to be considered. If \bar{D} has entries less than 0 and greater than L at the same time, we must fall back on theorem A.4, i.e. search through the possible combinations of D_i at the boundaries.

In the illustrating example of figure 1, when $\Gamma=1.5$, D_i obtained following step (i) in theorem A.6 are feasible. Hence, this single step finds the optimum D . When $\Gamma=2.9$, D_3 obtained following step (ii) is larger than 1, while D_1 and D_2 are smaller than 1. This satisfies the condition of step (iii). Hence, by fixing D_3 at 1 and following the calculation in theorem A.4, we find the optimum D^* .

We have discussed designing a matrix D to minimize the dominant eigenvalue of $D+G$ subject to constraints. This design allows us, for example, to allocate limited repair resources to an automaton network to best fight against the spread of a virus. It is instructive to study the structure of the optimizing $D=D^*$, and hence the structure of D^*+G , from the viewpoint of this application.

The structure of the optimizing D is highly dependent on the structure of the matrix G , which describes the connection topology of the network. The theorems give us the insight that, for a symmetric G , the matrix D should be chosen to best equalize the row sums of $D+G$ within the permitted constraints. In terms of resource allocation, this means that placing the most resources at the nodes that have strong connections best prevents virus spread. This makes sense since these nodes have the strongest potential to spread the virus throughout the network if they are infected and similarly to heal the network when they are healthy. Eliminating viruses at these nodes as soon as possible can quickly quench the spread. In case the individual constraints prevent placing enough resources at a node, nearby nodes are provided with extra resources to prevent the spread.

It is worth noting that this design is suitable for repair resource allocation before the breakout of a virus or in real time during a virus. In other words, this design is robust to the initial location of the virus. This is useful even when real-time allocation of resources after the start of an epidemic is not possible, or when it is hard to locate and respond to infected nodes throughout the network in an epidemic. When the initially affected nodes are known, the design can be improved further using this additional information. We leave this improvement to later work.

References

- Albert, R. & Barabási, A.-L. 2000 Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* **85**, 5234–5237. (doi:10.1103/PhysRevLett.85.5234)
- Anderson, P. M. & Fouad, A. A. 2003 *Power system control and stability power engineering series*, 2nd edn. Hoboken, NJ: IEEE Press.
- Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans dynamics and control*. Oxford, UK: Oxford University Press.

- Bak, P., Tang, C. & Wiesenfeld, K. 1987 Self-organized criticality: an explanation of $1/f$ noise. *Phys. Rev. Lett.* **59**, 381–384. (doi:10.1103/PhysRevLett.59.381)
- Bayen, A. M., Raffard, R. L. & Tomlin, C. J. 2004 Adjoint-based constrained control of Eulerian transportation networks: application to air traffic control. In *Proc. 2004 American Control Conf., Boston, MA*.
- Berman, A. & Plemmons, R. J. 1994 *Nonnegative matrices in the mathematical sciences*. Philadelphia, PA: Society of Industrial and Applied Mathematics Press.
- Bertsekas, D. P. 1995 *Nonlinear programming*, 2nd edn. Belmont, MA: Athena Scientific.
- Carlson, J. M. & Doyle, J. 1999 Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* **60**, 1412–1427. (doi:10.1103/PhysRevE.60.1412)
- Carreras, B. A., Lynch, V. E., Dobson, I. & Newman, D. E. 2002 Dynamics, criticality and self-organization in a model for blackouts in power transmission systems. In *Proc. Hawaii Int. Conf. on Systems Sciences, Hawaii, January 2002*.
- Chung, F. R. K. 1994 *Spectral graph theory*, published for the Conference Board of the Mathematical Sciences by the American Mathematical Society, no. 92.
- Costa, M., Crowcroft, J., Castro, M. & Rowstron, A. 2004 Can we contain internet worms. Microsoft Research technical report MSR-TR-2004-83.
- de Menezes, M. A. & Barabási, A.-L. 2004 Fluctuations in network dynamics. *Phys. Rev. Lett.* **92**, 028 701. (doi:10.1103/PhysRevLett.92.028701)
- Diekmann O. & Heesterbeek J. A. P. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Wiley series in mathematical and computational biology. New York, NY: Wiley.
- Farley, T., Foster, J. D., Hoang, T. & Lee, K. K. 2001 A time-based approach to metering arrival traffic to Philadelphia. In *Proc. First FIAA Aircraft Technology, Integration, and Operations Forum, Los Angeles, California*.
- Fisher, M. E. & Fuller, A. T. 1958 On the convergence of matrices and the convergence of linear iterative processes. *Proc. Camb. Philos. Soc.* **45**, 417–425.
- Gallager, R. G. 1996 *Discrete stochastic processes*. Norwell, MA: Kluwer Academic Publishers.
- Gevros, P. & Crowcroft, G. 2004 Distributed resource management with heterogeneous linear controls. *Comput. Netw. Int. J. Comput. Telecommun. Netw.* **45**, 835–858.
- Greenbaum, A. 1997 *Iterative methods for solving linear systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM) Press.
- Hekmat, R. 2006 *Ad-hoc networks: fundamental properties and network topologies*. Cambridge, MA: Springer.
- Hershkovitz, D. 1992 Recent directions in matrix stability. *Linear Algebra Appl.* **171**, 161–186. (doi:10.1016/0024-3795(92)90257-B)
- Hoang, T., Farley, T. & Davis, T. 2002 The multi-center TMA system architecture and its impact on inter-facility collaboration. In *Proc. AIAA Aircraft Technology, Integration and Operations (ATIO) Conference, Los Angeles, CA*.
- Kelly, F. 2003 Fairness and stability of end-to-end congestion control. *Eur. J. Control* **9**, 159–176. (doi:10.3166/ejc.9.159-176)
- Levins, R. 1969 Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15**, 237–240.
- Newman, M., Watts, D. & Barabasi, A.-L. 2006 *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.
- Olfati-Saber R., Fax J. A. & Murray R. M. 2007 Consensus and cooperation in networked multi-agent systems. In *Proc. IEEE, January 2007*.
- Perez-Arriaga, L. J., Verghese, G. C., Pagola, F. L., Sancha, J. L. & Schweppe, F. C. 1990 Developments in selective modal analysis of small-signal stability in electric power systems. *Automatica* **26**, 215–231. (doi:10.1016/0005-1098(90)90117-Z)
- Riley, S. et al. 2003 Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966. (doi:10.1126/science.1086478)

- Roy, S. & Saberi, A. 2007 Scaling: a canonical design problem for networks. *Int. J. Control* **80**, 1342–1353. (doi:10.1080/00207170701355069)
- Roy, S., Saberi, A. & Herlugson, K. 2006 *Formation and alignment of distributed sensing agents with double-integrator dynamics*. IEEE Press monograph on sensor network operations.
- Roy, S., Wan, Y. & Saberi, A. Submitted. Majorizations for the dominant eigenvector of a nonnegative matrix.
- Siljak, D. 1994 *Decentralized control of complex systems*. Boston, MA: Academic Press.
- Stewart, G. W. 1974 *Introduction to matrix computations*. New York, NY: Academic Press.
- Strogatz, S. H. 1993 Coupled oscillators and biological synchronization. *Sci. Am.* **269**, 102–109.
- Wan, Y. & Roy, S. In press. A scalable methodology for evaluating and designing coordinated air traffic flow management strategies under uncertainty. *IEEE Trans. Intell. Transport. Syst.*
- Wan, Y., Roy, S. & Saberi, A. Submitted. Designing spatially-heterogeneous strategies for control of virus spread.
- Wang, S. & Davison, E. J. 1973 On the stabilization of decentralized control systems. *IEEE Trans. Automat. Contr.* **AC-18**, 473–478. (doi:10.1109/TAC.1973.1100362)
- Wang, C., Knight, J. C. & Elder, M. C. 2000 On computer viral infection and the effect of immunization. In *Proc. 16th Annual Computer Security Applications Conference, New Orleans, Louisiana*.
- Wang, H., Xie, H., Qiu, L., Yang, Y., Zhang, Z. & Greenberg, A. 2006 COPE: traffic engineering in dynamic networks. In *Proc. SIGCOMM 2006, Pisa, Italy*.
- Watts, D. J. 2003 *Small worlds: the dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press.
- Watts, D. J. 2004 *Six degrees: the science of a connected age*. New York, NY: W. W. Norton and Co.